A COHORT MODEL FOR PREDICTING RETENTION OF REGULAR MARINE CORPS OFFICERS

Carlos Kay McAfee

United States
Naval Postgraduate School

THESIS

A COHORT MODEL FOR PREDICTING
RETENTION OF REGULAR MARINE CORPS OFFICERS

by

Carlos Kay McAfee

April 1970

A Cohort Model for Predicting

Retention of Regular Marine Corps Officers

by

Carlos Kay McAfee
Major United States Marine Corps
B.S., United States Naval Academy, 1955

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
April 1970

ABSTRACT

A model for predicting retention of regular Marine Corps officers is formulated. Using data from the three cohorts 1956 - 58, a statistical test is used to show that the lifetimes on active duty of members of each cohort follow the same distribution. Unknown parameters for this distribution are estimated from the data. Retention figures for various lengths of service of the cohorts 1959 - 63 are calculated with the model and compared with actual data. Confidence intervals for the predictions are given. Use of the model and follow-on studies are suggested and examples given.

TABLE OF CONTENTS

I. SUMMARY

A. INTRODUCTION

The need for predicting and controlling the number of officers on active duty in the Marine Corps is a very real and continuing problem. Since overall officer strength is established by law, personnel planners must determine the number of officers to be introduced into the system so that future needs will be met without exceeding prescribed limits.

Marine Corps officers come on active duty with either a regular or reserve commission. Officers with regular commissions have a specified obligated period of service, after which they may either resign their commission or remain on active duty as they so desire. On the other hand, reserve officers have three alternatives. During their initial obligated period of service they may be given the opportunity to apply for a regular commission; if they do so and are accepted they can then stay on active duty or resign (after their obligated time) as they so desire. Their other alternatives are to leave active duty after completing their obligated period or to apply for additional time on active duty as a reserve, subject to approval of the Commandant.

The fact that regular officers have the freedom to choose when they will leave active duty raises the question of how can one predict the number of officers that will be on active duty some time in the future. This problem of predicting future group sizes of regular officers is the motivation for this study.

B. PURPOSE

This study was undertaken to investigate attrition and retention

rates of different year-groups (henceforth called cohorts) of regular Marine Corps officers. Cohorts studied consisted of all officers receiving commissions as regular Marine Corps Second Lieutenants during a calendar year.

Data was taken for three cohorts, 1956 - 58. After computing the retention rates realized from these groups, the stationarity properties were investigated.

Let $p_i(jk)$ = P [Member of cohort 19jk is on active duty i years after commissioning].

If cohorts behave essentially the same for different year groups, then $p_i(jk) = p_i$ for all jk. That is, the p_i are stationary over time and independent of the cohort initial size. This hypothesis is tested in Section III B. Results of the test lead to the formulation of a prediction model, and predictions for year groups 1959 - 63 are compared with actual figures in Section III D.

C. RESULTS

Analysis of cohorts 1956 - 58 indicate that attrition from these groups can be considered stationary at least through 1968, the last year that data was available at the time of the study. A prediction model was formulated based on the overall yearly retention realized from the total membership of these three cohorts. For example, p_i is computed by dividing the total number of officers from cohorts 1956 - 58 still on active duty i years after commissioning by the total initial size of the three groups. Using these p_i predictions are made of the number remaining on active duty in various years from cohorts 1959 - 63. These predictions agree well with actual figures for these cohorts.

This study does not attempt to predict retention by rank or promotion frequency, nor is the reason for officers leaving active duty considered.

Section IV contains a discussion of the similarities and differences between the cohort model and Markov Chain Models which have been discussed widely in the literature.

In Section V A, it is suggested that the methodology used to develop the model can be employed to develop similar models for other cohorts such as reserve officers, aviators, etc. After developing such models, a cost model could be developed to compare the procurement costs per expected year of service for various cohorts.

It is suggested in Section V B that an application of the model might be to assist personnel planners in determining the number of reserve officers from a given year group to be given regular commissions.

An example is given in Section V B which illustrates how this model, in conjunction with a similar one for reserve officers, might be used to assist in determining and controlling future numbers of Marine Corps officers.

II. DATA COLLECTION

A. COLLECTION PROCEDURES

Cohort size for a given year was determined from applicable sections of the following year's Lineal List [Ref. 1]. For example, the group of 1956 consists of all officers receiving regular Second Lieutenant commissions during the calendar year 1956, according to the 1957 Lineal List. Subsequent cohort size was established by counting those officers of an initial group that were still on active duty according to the appropriate Lineal List. An officer was considered only to be on active duty or not. No attempt was made to differentiate between different reasons for leaving, and no attempt was made to consider promotions or occupational specialty.

B. DATA BASE

Table I below lists the number and percentage of each cohort on active duty in succeeding calendar years after their initial commission. Two graphs of this data are given. Each is plotted using semi-logarithmic scale. In Graph I the plot shows the percentage remaining in relation to the year after commissioning. In Graph II the percentage remaining is plotted against a common origin of clock time starting in year 1956. A straight line on these graphs would result from attrition at a constant rate. Concave and convex segments of the plots indicate increasing and decreasing attrition rates, respectively.

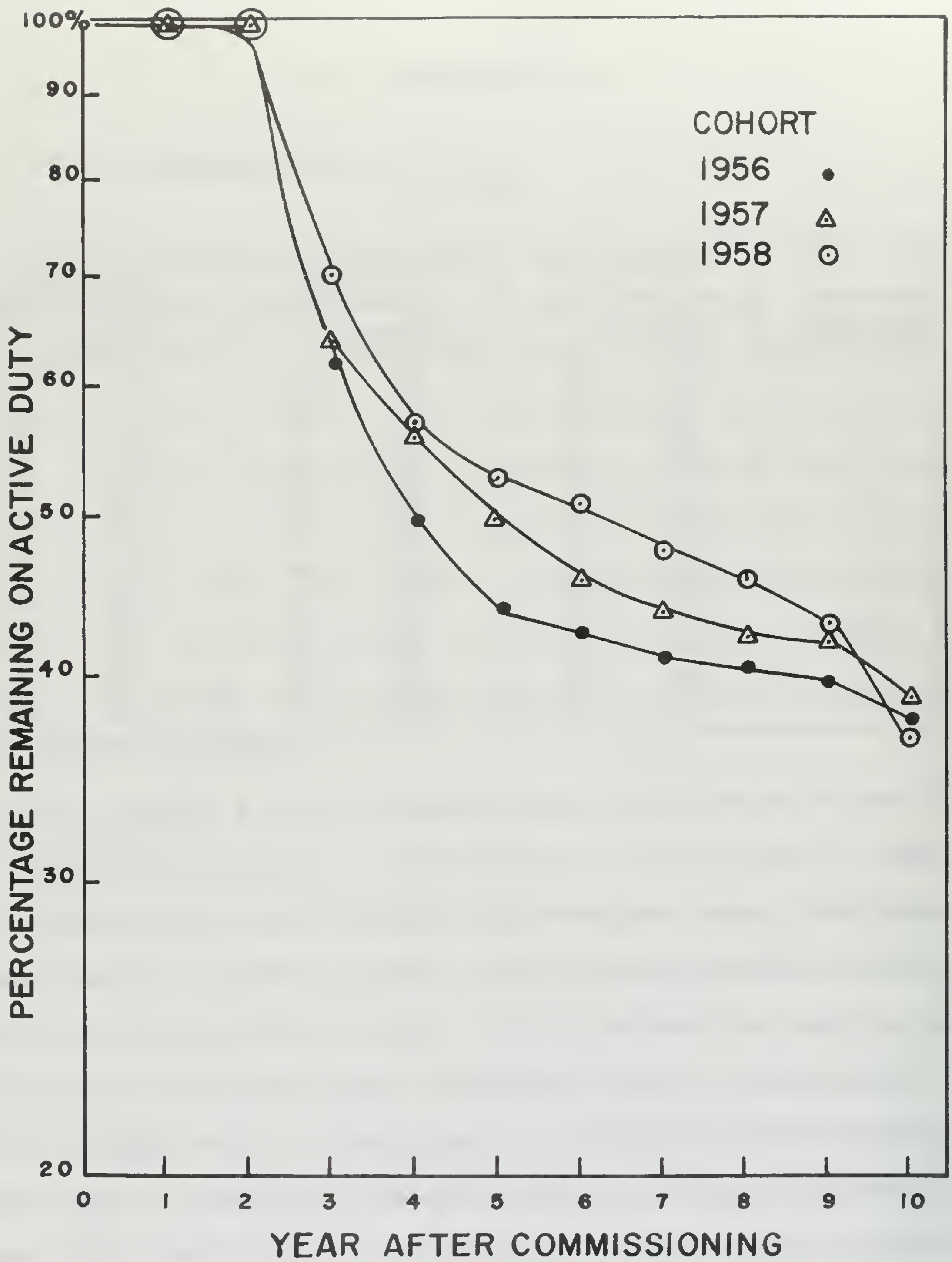
Some important characteristics of the data are quite apparent in these graphs. There is an initial period of nearly 100% retention which

TABLE I

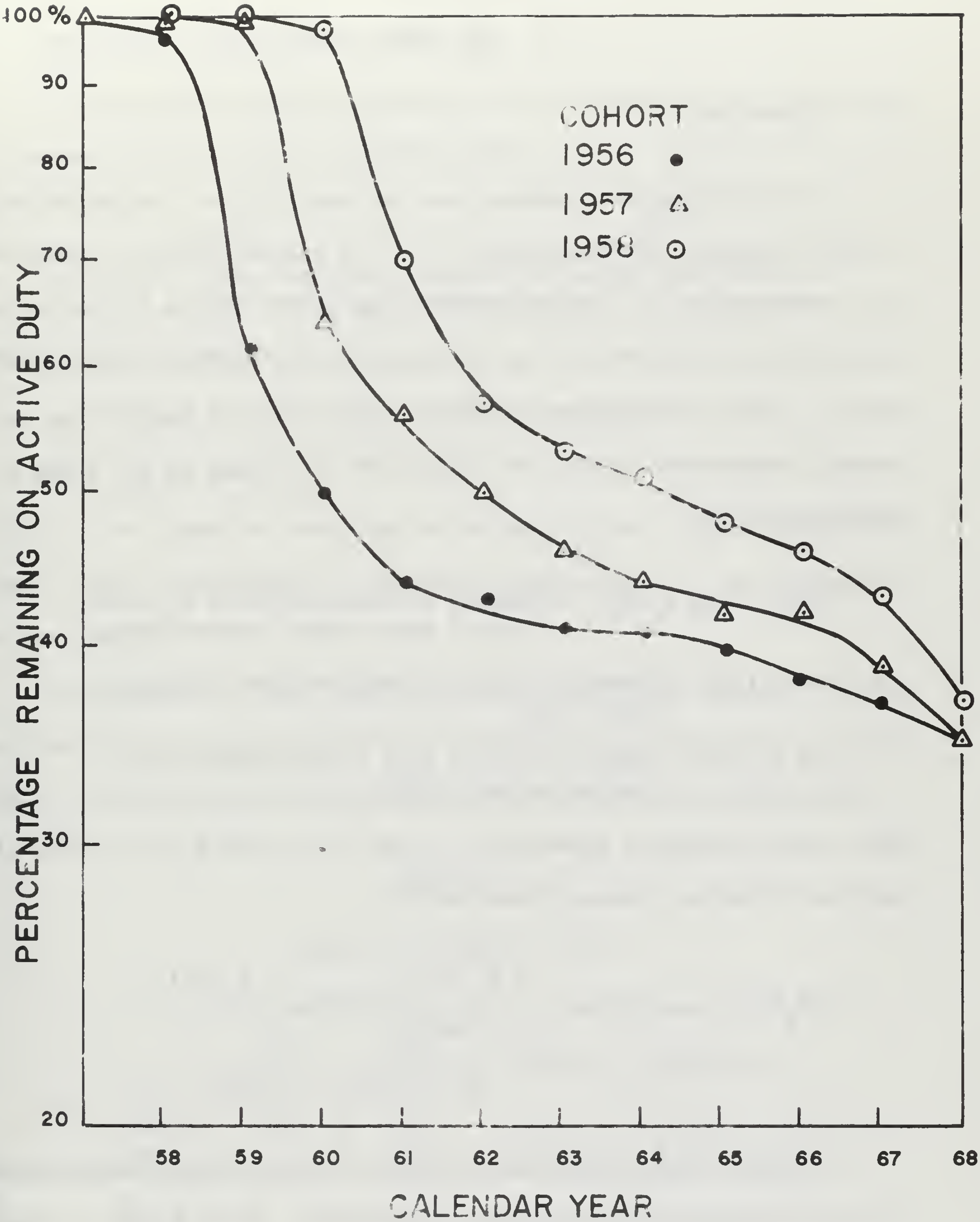
Data For Regular Marine Corps Officer Cohorts
1956 - 58

Year After Commission	1956 Cohort On Active Duty Number	%	1957 Cohort On Active Duty Number	%	1958 Cohort On Active Duty Number	%
0	414	100	339	100	369	100
1	410	99	337	99	367	99
2	408	99	334	99	362	98
3	256	62	218	64	257	70
4	206	50	190	56	211	57
5	182	44	168	50	196	53
6	176	43	156	46	187	51
7	171	41	150	44	178	48
8	168	41	144	42	168	46
9	165	40	141	42	157	43
10	156	38	133	39	137	37
11	153	37	120	35	NA	NA
12	143	35	NA	NA	NA	NA

is due to the obligation of each officer to serve a minimum of three years on active duty after commissioning. Following this obligation period both graphs show essentially similar slopes for each cohort, a generally decreasing attrition rate. Graph II shows a tendency toward an increased attrition rate for all groups in 1966 and 1967. It is not the purpose of this paper to analyze or explain these characteristics, and no attempt is made to do so. The graphs do however suggest a model, in that they appear to be plots of three realizations of the same, but unknown, random process. Formulation of the model and hypothesis testing are discussed in the next section.



GRAPH I. PERCENTAGE REMAINING ON ACTIVE DUTY RELATIVE TO YEAR AFTER COMMISSIONING.



GRAPH II PERCENTAGE REMAINING ON ACTIVE DUTY BY CALENDAR YEAR

III. THE COHORT MODEL

A. ASSUMPTIONS

The following basic assumptions are made in the formulation of the officer retention prediction model. It is assumed that all individuals act independently of each other with regards to leaving or staying on active duty. Each officer can be considered to undergo a random walk where p_i is the probability he stays in the system i years after entrance, independent of when he joined and the number in the group in which he joined.

Recall that $p_i(jk)$ = P[Member of cohort 19jk is on active duty i years after commissioning].

Let $X_o(jk)$ = Number of officers commissioned in calendar year 19jk,

and $X_i(jk)$ = Number on active duty i years after comissioning in year 19jk.

Then, under the above assumptions, $p_i(jk) = p_i$, and $X_i(jk)$ given $X_o(jk)$, has the following binomial distribution;

$$P[X_i(jk) = n \mid X_o(jk)] = \binom{X_o(jk)}{n} p_i^{\,n} (1-p_i)^{X_o(jk)-n} \quad ,$$

$$n = 0,1,2,\ldots,X_o(jk).$$

The theory that data from the 1956 - 58 cohorts are realizations of this stochastic process must now be tested. Since $X_o(jk)$ is large (approximately 400) we can use the well known homogeniety test (see, for example, Guenther [Ref. 2]). This test establishes whether we can accept our theory based on the comparison of expected and observed cohort sizes from year to year. The test and its results are discussed below.

B. HYPOTHESIS TESTING

To verify the foregoing assumption of stationarity, a hypothesis of homogeniety, as found in Guenther [Ref. 2], is tested for cohorts 1956 - 58.

Let $r_i(jk)$ = The expected fraction of $X_o(jk)$ on active duty i years after commissioning.

Now test the hypothesis:

H_o: $r_i(jk) = r_i$, $i = 3,4,...9$; $jk = 56,57,58$.

H_1: $r_i(jk) \neq r_i$.

That is, the fraction remaining on active duty for a given number of years after commissioning is tested to determine if it is independent of the cohort.

The homogeniety test compares realizations with expected values, and can best be illustrated with a contingency table, Table II, where the O_{ij} are observed cohort sizes a given year after commissioning.

TABLE II

Sample Contingency Table

	Cohorts 1956	1957	1958	Totals
On Active Duty	O_{11}	O_{12}	O_{13}	$\sum_j O_{1j}$
Not on Active Duty	O_{21}	O_{22}	O_{23}	$\sum_j O_{2j}$
Totals	$\sum_i O_{i1}$	$\sum_i O_{i2}$	$\sum_i O_{i3}$	$\sum_i \sum_j O_{ij}$

Using the elements of the contingency table, expected cohort sizes, E_{ij}, are computed as follows:

$$E_{11} = \frac{(\sum_j O_{1j})\ (\sum_i O_{i1})}{\sum_i \sum_j O_{ij}}$$

$$E_{12} = \frac{(\sum_j O_{1j})\ (\sum_i O_{i2})}{\sum_i \sum_j O_{ij}}$$

$$\begin{array}{cc} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{array}$$

$$E_{23} = \frac{(\sum_j O_{2j})\ (\sum_i O_{i3})}{\sum_i \sum_j O_{ij}}$$

The test value is then computed by taking

$$\sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} .$$

This result is compared to the appropriate value from the chi-square table. If the computed value is less than, or equal to, the tabled value, the hypothesis is accepted.

An example of the computations done in this analysis is shown below. Observed data for the third year after commissioning is listed in Table III. Expected values for the third year after commissioning were computed as follows:

TABLE III

Observed Values for the Third Year After Commissioning

	Cohort 1956	1957	1958	Totals
On Active Duty	256	218	257	731
Not On Active Duty	158	121	112	391
Totals	414	339	369	1122

$$E_{11} = \frac{414 \times 731}{1122} = 270$$

$$E_{12} = \frac{339 \times 731}{1122} = 221$$

$$E_{13} = \frac{369 \times 731}{1122} = 240$$

$$E_{21} = \frac{414 \times 391}{1122} = 144$$

$$E_{22} = \frac{339 \times 391}{1122} = 118$$

$$E_{23} = \frac{369 \times 391}{1122} = 129$$

The test value is computed as:

$$\text{Test Value} = \frac{(257 - 270)^2}{270} + \frac{(218 - 221)^2}{221} + \frac{(257 - 240)^2}{240} +$$

$$\frac{(158 - 144)^2}{144} + \frac{(121 - 118)^2}{118} + \frac{(112 - 129)^2}{129} = 4.56.$$

The value from the chi-square table, at a 5% level, with two degrees of freedom is 5.99. Therefore the hypothesis of homogeniety for year three is accepted. Similar computations were done for years four through nine and the results are shown in Table IV. Analysis indicates that for year

nine the hypothesis should be rejected, however the overall hypothesis of homogeniety can not be rejected by this analysis.

TABLE IV

Homogeniety Test Results

Year	Test Value	Tabled Value
3	4.56	5.99
4	4.12	5.99
5	3.59	5.99
6	3.26	5.99
7	1.18	5.99
8	4.11	5.99
9	6.71	5.99

C. PARAMETER ESTIMATION

The parameters, p_i, for the model are estimated using the average outcome of cohorts 1956 - 58. For a given year, say year i, after commissioning p_i is determined as follows:

$$p_i = \frac{X_i\ (56) + X_i\ (57) + X_i\ (58)}{X_o\ (56) + X_o\ (57) + X_o\ (58)} .$$

Using this procedure, the p_i for years three through nine were computed and are listed in Table V.

TABLE V

Model Parameters, p_i

Year	p_i
3	.65
4	.54
5	.49
6	.47
7	.45
8	.43
9	.41

These probabilities are used to compute expected sizes and confidence limits for other cohorts. For example, the expected number remaining in cohort 1961, i years after commissioning is:

$$E_i\ (61) = [X_o\ (61)]\ p_i\ ,\ i = 3,4,\ldots,9.$$

The confidence interval (CI) for this predicted value is computed as follows:

$$CI_i\ (61) = [E_i\ (61) - 2\sqrt{Var_i\ (61)}\ ,\ E_i\ (61) + 2\sqrt{Var_i\ (61)}],$$

where $Var_i\ (61) = [X_o\ (61)]\ p_i\ (1-p_i)\ ,\ i = 3,4,\ldots,9.$

Since we have binomial distributions with large n, using the Central Limit Theorem, we can assume normality. Therefore the use of two standard deviations gives a 95% confidence interval.

D. PREDICTION RESULTS AND COMPARISON WITH REAL DATA

Using the procedures described in Section II, selected data was collected on cohorts 1959-63. In Tables VI and VII, this data is compared with predictions made using the model with the parameters in Table V. Each of the actual cohort sizes was within the confidence interval with but one exception, year four of cohort 1960.

TABLE VI

Actual Numbers, Predicted Numbers and Confidence Intervals For Cohort 1960

Year After Commissioning	Actual Numbers	Predicted Numbers	95% Confidence Intervals
3	240	229	(211,247)
4	212	191	(173,209)
5	190	173	(155,191)
6	172	166	(148,184)
7	161	159	(141,177)
8	144	152	(134,170)

TABLE VII

Actual Numbers, Predicted Numbers and Confidence Intervals For Selected Cohorts on Active Duty in 1968.

Cohort	Actual Numbers	Predicted Numbers	95% Confidence Intervals
1959	139	147	(129,165)
1960	144	152	(134,170)
1961	141	141	(123,159)
1962	173	163	(145,181)
1963	138	141	(125,157)

IV. MARKOVIAN PROPERTIES OF THE COHORT MODEL

A significant number of recent models of social behavior use Markov chain theory. Several such models are discussed by Bartholomew and Thonstad [Refs. 3 and 4]. Because the use of Markov chain theory is so widespread in the literature, it is appropriate to point out the similarities and differences between the cohort model developed in this study and the Markov Chain Models.

Let $Y(jk)$ be the expected total number, from all cohorts considered, on active duty in year 19jk.

Using the notation shown in Section III,

$$Y(jk) = X_o(jk) + X_o(jk-1)p_1 + X_o(jk-2)p_2 + \dots$$

Now let $C(jk)$ = The fraction of $Y(jk)$ who remain on active duty in year 19jk + 1.

$$\text{So } C(jk) = \frac{Y(jk+1) - X_o(jk+1)}{Y(jk)}$$

$$= \frac{X_o(jk)p_1 + X_o(jk-1)p_2 + X_o(jk-2)p_3 + \dots}{X_o(jk) + X_o(jk-1)p_1 + X_o(jk-2)p_2 + \dots} \quad . \qquad (1)$$

$C(jk)$ can be interpreted as the probability of remaining on active duty in year (jk+1) given you are on active duty in year (jk). Clearly this is a function of all previous cohort sizes and in general is a function of (jk).

Now assume a constant attrition rate, or equivalently that the curves in Graphs I and II are straight lines. Then define:

$$p = P[\text{Person is on active duty in year } t \mid \text{on active duty in year } t-1.] \qquad (2)$$

Clearly $p_t = p^t$ for all t.

Now equation (1) becomes

$$C(jk) = \frac{X_o(jk)p + X_o(jk-1)p^2 + X_o(jk-2)p^3 + \ldots}{X_o(jk) + X_o(jk-1)p + X_o(jk-2)p^2 + \ldots}$$

Multiplying top and bottom by p, and cancelling, results in $C(t) = p$ for all t; that is, with a constant attrition rate the expected fraction remaining on active duty from year to year is constant. In this case attrition from a cross-section depends only on the size of the cross-section and not on the fraction of each cohort in the cross-section. Thus the Markov property holds and the cohort model is equivalent to a Markov Model, where we define:

State 1: On active duty.

State 2: Not on active duty.

The transition probability matrix is

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} p & 1-p \\ 0 & 1 \end{bmatrix} \end{array}, \text{ where } p \text{ is defined in (2).}$$

It is evident from Graphs I and II that the attrition rate is not constant for Marine Corps regular officers.

Let us assume that, rather than a constant attrition rate, all cohorts are of the same initial size, that is $X_o(jk) = X$ for all jk. Now equation (1) becomes:

$$C(jk) = \frac{X(p_1 + p_2 + p_3 + \ldots p_j + \ldots)}{X(1 + p_1 + p_2 + \ldots p_j + \ldots)}$$

$$= \frac{\sum_{j=1}^{\infty} p_j}{1 + \sum_{j=1}^{\infty} p_j} .$$

C(jk) in this case is again a constant, independent of time. In general

the overall attrition rate from a given cross-section does depend on the fraction of the cross-section which comes from a cohort of a given age. But if all cohorts have the same initial size the fraction of the cohort of age t, say, is the same in every cross-section. This leads to great stability in the overall attrition rates in succeeding years.

We conclude that if cohort sizes remain nearly constant from year to year, then a Markov-type model will probably give good predictions. However, since initial cohort sizes in future years are to be thought of as control variables, the cohort model should describe attrition phenomena more accurately than the Markov-type Model. Indeed the figures in Table I indicate that in the past cohort sizes have differed significantly.

V. CONCLUSIONS

The conclusion reached from this study is that retention rates from different cohorts can be considered stationary over reasonable time periods independent of time and of group size. The groups looked at were commissioned up to seven years apart. Engagement in the Viet Nam conflict occurred at different times relative to each cohort's entry into the Marine Corps. Years between promotions varied somewhat and even obligated service was not of the same duration for all groups. However with all these effects taking place, behavior was essentially stationary and the model provides statistically valid predictions for all cohorts considered.

A. FURTHER STUDY

The methodology developed in this study can be applied to other groups of interest. Possible cohorts might be: aviators, Naval Academy graduates, officers that were previously enlisted Marines, and regular officers that were initially commissioned as reserves.

A natural follow-on study would be to develop a cost model for cohorts of interest. By using the average cost of obtaining an officer of a particular group and his expected length of service, as computed in his cohort analysis, an average cost per expected year of service could be compared for different cohorts to assist in determining the optimal means of officer procurement and retention.

B. USE OF THE STUDY

The ability to predict future sizes of incoming officer cohorts undoubtedly has several immediate applications in personnel planning. One might be to assist in determining the number of reserve officers from a given year group to receive regular commissions. Using this model the planner can estimate the number of regular officers that will be on active duty at some time in the future, comparison of this figure with anticipated requirements will give an indication of how many reserves will be needed to meet needs.

Here is a simple example of how this model and methodology can be used to assist in officer planning. Assume that as suggested in Section IV A above, a similar model is developed for all other officer cohorts, i.e., for those not initially commissioned as regular officers, and that the new model has stationarity characteristics similar to the model developed in this study.

Let: p_i' = P[Officer receiving a reserve commission is on active duty i years after commission.]

$X_o'(jk)$ = Number of officers entering active duty with reserve commissions during calendar year jk.

$X_i'(jk)$ = Number of these on active duty i years after commission.

Now by combining the two models, we can look at expected retention over a period of time, say five years for this example. This gives us the following:

$$\begin{bmatrix}\text{Expected number of} \\ \text{officers, commissioned} \\ \text{during 1970-74, on} \\ \text{active duty in 1979.}\end{bmatrix} = X_o(70)p_9 + X_o'(70)p_9' + X_o(71)p_8 + X_o'(71)p_8' \; . \; . \; . + X_o(74)p_5 + X_o'(74)p_5'.$$

With this function we can address several different questions concerning officer retention. Three possible areas of interest are:

1. If the $X_o(jk)$ and $X_o'(jk)$ are fixed, we can predict how many officers from these cohorts will be on active duty in 1979.

2. If some $X_o(jk)$ and $X_o'(jk)$ are fixed and some variable, to meet a desired total from these cohorts in 1979, we can determine optimal sizes for the variable inputs.

3. If all $X_o(jk)$ and $X_o'(jk)$ are fixed and we have upper and lower limits on the total number that we want in 1979, then steps can be taken to increase or decrease attrition rates as appropriate.

LIST OF REFERENCES

1. NAVMC P-1005-PD, Combined Lineal List of Officers on Active Duty in the Marine Corps, 1957 - 1969.

2. Guenther, W. C., Concepts of Statistical Inference, p. 187 - 195, McGraw-Hill, 1965.

3. Bartholomew, D. J., Stochastic Models for Social Processes, John Wiley and Sons, 1967.

4. Thonstad, T., Education and Manpower, University of Toronto Press, Toronto, 1968.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Asst. Professor K. T. Marshall, Code 55Mt Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	10
4.	Major Carlos K. McAfee, USMC 4313 N. St. Clair Oklahoma City, Oklahoma 73112	2
5.	Commandant of the Marine Corps (Code A03C) Headquarters, U. S. Marine Corps Washington, D. C. 20380	1
6.	James Carson Breckinridge Library Marine Corps Development and Educational Command Quantico, Virginia 22134	1
7.	Dr. R. M. Oliver Administrative Studies Project Office of Institutional Research Room 210 - T8 University of California Berkeley, California 94720	1
8.	Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	1

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3 REPORT TITLE

A Cohort Model for Predicting Retention of Regular Marine Corps Officers

4 DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Master's Thesis; April 1970

5 AUTHOR(S) (First name, middle initial, last name)

Carlos Kay McAfee

6 REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1970	27	4

8a. CONTRACT OR GRANT NO. b. PROJECT NO c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 DISTRIBUTION STATEMENT

This document has been approved for public release and sale, its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

13 ABSTRACT

A model for predicting retention of regular Marine Corps officers is formulated. Using data from the three cohorts 1956 - 58, a statistical test is used to show that the lifetimes on active duty of members of each cohort follow the same distribution. Unknown parameters for this distribution are estimated from the data. Retention figures for various lengths of service of the cohorts 1959 - 63 are calculated with the model and compared with actual data. Confidence intervals for the predictions are given. Use of the model and follow-on studies are suggested and examples given.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Marine Corps Officers						
Manpower						
Personnel						
Prediction						
Markov Chains						
Cohort Models						